Article | Received 6 September 2025; Revised 21 October 2025; Accepted 21 November 2025; Published 26 November 2025
https://doi.org/10.55092/rl20250010

Social GAN guided multimodal trajectory prediction and MPC for autonomous driving

Zihan Jiang[1,†], Haochen Zhu[2,†], Ruonan Liu[2,*], Xiuzhong Hu[2] and Weidong Zhang[2,3,*]

[1] Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai 200000, China
[2] School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai 200240, China
[3] Henan Key Laboratory of Big Data Analysis and Processing, School of Computer and Information Engineering, Henan University, Kaifeng 475001, China

† These authors contributed equally to this work.

* Correspondence authors; E-mails: ruonan.liu@sjtu.edu.cn (R.L.); wdzhang@sjtu.edu.cn (W.Z.).

Highlights:

- Novel Integration: An end-to-end framework that fuses Social GAN's multimodal pedestrian predictions with a constraint-aware MPC planner for the first time.
- Zero Collisions: Achieves perfect safety (zero violations) by reasoning over diverse future intents in dense crowds.
- Minimal Trade-off: Maintains real-time performance and passenger comfort with only marginal travel time and control effort increases.

Abstract: Navigating autonomously in dense, dynamic environments requires autonomous vehicles to reason about multiple, socially plausible futures of surrounding pedestrians while ensuring safe and feasible motion. To bridge the gap between multimodal trajectory prediction and safety-critical control, this paper introduces a novel end-to-end prediction-to-control pipeline that seamlessly integrates socially consistent, multimodal trajectory proposals from a Social Generative Adversarial Network (Social GAN) directly into a constraint-aware Model Predictive Control (MPC) planner. Unlike approaches that compress predictions into a single mode or use them only as soft costs, our method explicitly incorporates the diverse trajectory hypotheses generated by Social GAN as time-varying dynamic obstacles within the MPC's optimization problem. This preserves the intent diversity crucial for negotiation in crowded spaces and allows the controller to plan for a spectrum of possible outcomes proactively. Evaluated on the ETH/UCY datasets and in dynamic crowd simulations, the integrated Social GAN+MPC controller achieves zero safety violations and maintains a larger average pedestrian clearance (0.94 m) compared to a reactive baseline, at the cost of only a marginal increase in travel time and a modest reduction in path efficiency due to risk-averse maneuvering. The increase in mean acceleration ($0.57\ \mathrm{m/s^2}$) and jerk ($1.34\ \mathrm{m/s^3}$) remains within acceptable comfort bounds, demonstrating that the framework delivers

proactive safety without compromising real-time feasibility. This work demonstrates that coupling calibrated, intent-diverse predictions with a robust optimizer provides a practical and effective pathway toward zero-collision autonomous driving.

Keywords: pedestrian trajectory prediction; safety-aware motion planning; Social GAN; model predictive control

1. Introduction

The rapid advancement of autonomous driving technology has the potential to transform transportation systems, offering increased safety, efficiency, and accessibility [1,2]. Autonomous vehicles (AVs) are expected to reduce human error, responsible for a significant percentage of traffic accidents [3,4]. They also promise to optimize traffic flow, reduce congestion, and provide greater mobility for individuals unable to drive due to age, disability, or other reasons. The ultimate goal of autonomous driving is to create vehicles that can navigate complex, dynamic environments with minimal or no human intervention, achieving full automation across a wide range of road conditions and traffic scenarios [5–7].

However, realizing this vision is a formidable challenge, particularly in urban environments where traffic is dense and unpredictable. Autonomous driving systems must be able to perceive their surroundings, make decisions in real-time, and execute motion plans that ensure the safety and comfort of passengers, pedestrians, and other vehicles [8,9]. To achieve this, two critical capabilities must be seamlessly integrated: accurate trajectory prediction of surrounding agents and robust, real-time motion planning that accounts for a wide range of possible future scenarios [10].

Predicting the behavior of pedestrians and other road users is inherently uncertain and multi-modal. People do not always follow predictable paths; they may hesitate, speed up, or change direction unexpectedly [11–13]. These behaviors must be considered when planning an autonomous vehicle's trajectory. Traditional motion planning techniques rely on simplified models of human behavior, which struggle to capture the complexity and variability inherent in pedestrian movement [14–16]. Human pedestrians often exhibit non-linear, unpredictable behaviors, such as hesitation, abrupt changes in speed, and interaction with other agents. These multi-modal behaviors pose significant challenges for conventional trajectory prediction models, which tend to provide deterministic predictions or fail to account for the inherent uncertainty of the environment [17–19]. The schematic diagram for the pedestrian trajectory prediction task is shown in Figure 1.

A persistent challenge in bridging prediction and planning lies in the trade-off between multimodality and computational tractability. On one hand, many practical planning systems, for the sake of real-time operation, compress the rich output of multimodal predictors into a single conservative prediction (e.g., a worst-case scenario) or a probabilistic occupancy map [20,21]. While efficient, this loss of intent diversity can lead to overly cautious or even infeasible plans when multiple distinct futures are equally probable. On the other hand, methods that naively consider a vast number of predicted samples to preserve diversity often incur prohibitive computational costs, making them unsuitable for the tight control loops required in autonomous driving [22]. This creates a critical gap for a framework that can preserve socially consistent multimodality for robust decision-making while remaining computationally lean enough for real-time application.

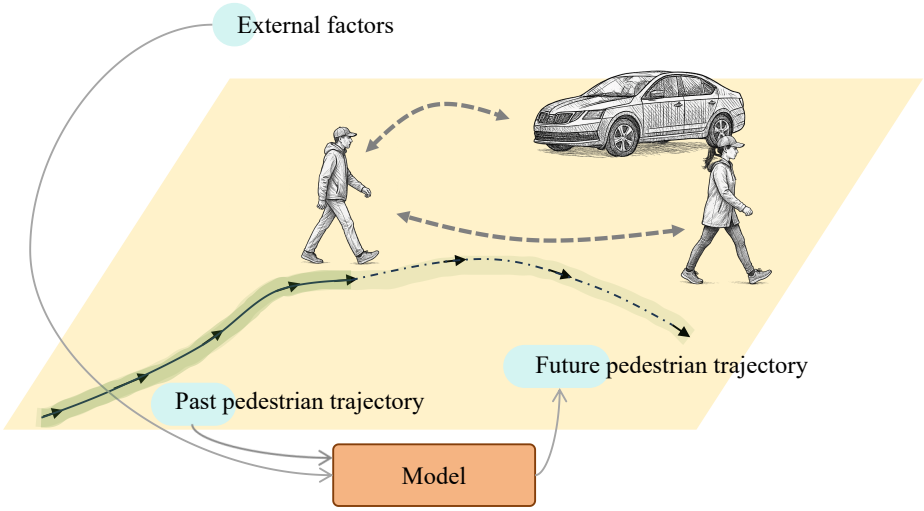

Figure 1. Pedestrian trajectory prediction task diagram.

Beyond this fundamental trade-off, a practical solution must also demonstrate real-world applicability. This entails not only meeting strict real-time computational budgets—a core requirement for any automotive-grade system—but also possessing a modular design that allows for integration into existing autonomy stacks without necessitating a complete architectural overhaul. A method that is either too slow for a high-frequency control loop or incompatible with modular perception-planning-actuation pipelines has limited deployment potential.

To address these challenges, this paper proposes a novel approach that combines the strengths of Social GAN [23], a generative adversarial network designed to model socially aware pedestrian trajectory prediction, with Model Predictive Control (MPC), a well-established optimization-based framework for real-time motion planning [24,25]. The proposed framework integrates prediction and control in a manner that allows for the simultaneous consideration of multiple plausible future trajectories and the optimization of vehicle motion in real-time. By incorporating the multi-modal predictions generated by Social GAN, the planner is able to account for a wider range of possible agent behaviors, thereby improving safety and robustness in dynamic environments.

Social GAN models pedestrian behavior as a function of individual intentions and social interactions, producing multiple future trajectory hypotheses representing different possible outcomes. This diversity of predictions is achieved through adversarial training and a variety of loss functions, which encourage the generation of trajectories that reflect the inherent variability of human motion [26–28]. This multi-modal approach ensures that the planner can respond to broader scenarios, reducing the risk of overly conservative or aggressive planning decisions [29,30]. On the control side, MPC generates optimal vehicle trajectories that respect dynamic constraints, including collision avoidance with pedestrians and other vehicles, and physical constraints such as speed and kinematic limitations [31–33]. The key advantage of MPC is its ability to incorporate prediction uncertainty into the planning process by treating the predicted trajectories as dynamic constraints that evolve [34,35]. This approach enables the planner to react in real-time to environmental changes while maintaining safe and smooth control.

Overall, the contributions of this paper are as follows:

(1) This paper provides a unified prediction-to-control pipeline, where Social GAN's multi-modal

trajectory predictions directly inform the MPC optimization process;

(2) We demonstrate the effectiveness of this integrated approach through extensive experimentation on standard pedestrian trajectory prediction benchmarks.

The remainder of this paper is structured as follows. Section 2 introduces relevant trajectory prediction and control models. Section 3 details the proposed method. Section 4 demonstrates the effectiveness of the proposed method through experimental results. Finally, Section 5 summarizes the content and outlines future research directions.

2. Related work

2.1. Trajectory prediction

Forecasting the motion of traffic participants is intrinsically multimodal and interaction-driven. Early deterministic regressors reduced trajectory forecasting to point estimates, which are brittle under aleatoric uncertainty and social intent variability. Generative approaches (e.g., GAN [36,37]/CVAE families [38,39]) reframed the task as learning a distribution over futures, encouraging sample diversity so that multiple socially plausible paths coexist rather than collapsing to a single mean. Social pooling and its successors (graph neural networks [40,41] and attention/transformer architectures [42,43]) made inter-agent coupling explicit, letting the predictor condition on neighbors' states and intents. Map-aware models further exploited lane topology, drivable areas, and semantic priors, aligning hypotheses to the scene structure and drastically reducing off-map artifacts.

Beyond GAN/CVAE, flow-based and diffusion-style predictors brought complementary strengths: normalizing flows provide exact likelihoods and efficient sampling, which benefits calibration and risk-aware downstream use, while diffusion models tend to produce broad, mode-covering samples with high geometric fidelity at the cost of higher compute. Mixture-of-experts and hierarchical decoders split long-horizon uncertainty into interpretable alternatives, improving coverage and controllability. Another thread reasons over occupancy or motion fields rather than per-agent polylines, which can better capture group phenomena but can be harder to feed into classical controllers.

A persistent challenge is the diversity–accuracy trade-off. Metrics such as ADE/FDE reward mode coverage, yet excessive diversity can dilute accuracy for the realized mode [44]; conversely, narrow distributions look good under average errors but fail catastrophically in rare yet safety-critical cases. Recent works therefore emphasize calibration (matching predicted and empirical frequencies), uncertainty disentanglement (aleatoric *vs.* epistemic), and evaluation under closed loop—measuring how forecasts perform when actually coupled to a planner, rather than in open-loop replay [45–47]. These lessons motivate treating predictors not as oracles but as proposal generators whose samples must be filtered and optimized by decision-making [48].

2.2. Control optimization beyond a single paradigm

MPC is widely used because it naturally accommodates constraints and receding-horizon re-optimization [49]. However, production-grade autonomy leverages a portfolio of control and planning paradigms. Search-/lattice-based planners efficiently enumerate kinematically feasible motions on structured roads,

while sampling-based controllers (e.g., CEM/MPPI) evaluate thousands of stochastic rollouts against learned or analytic costs to escape local minima [50]. Rule and behavior-based layers handle high-level intents, handing motion targets to a lower-level optimizer [51]. From the safety side, Control Barrier Functions (CBFs) [52], viability kernels, and reachability analysis provide certificates or conservative sets for collision avoidance.

Risk handling is increasingly distribution-aware: chance constraints trade constraint violation probability against performance [53]; CVaR and other coherent risk measures focus the controller on tail events [54]; distributionally robust optimization (DRO) hedges against model mis-specification by optimizing for the worst case within an ambiguity set [55]. Scenario-based MPC samples multiple futures (weather, intents) and optimizes feasible policies across those scenarios [56]. To meet real-time budgets, practical systems combine these ideas with engineering enablers—code-generated solvers, warm starts, parallel evaluation on accelerators. The limitation, however, is that many pipelines still rely on unimodal or heuristically pruned forecasts, which under-represent social multimodality and can yield brittle plans when the realized intent deviates from the expected one.

2.3. Learning and control integration

Moving beyond open-loop forecasting, prediction-aware planning pipelines explicitly feed distributions into the decision layer. Some works decouple prediction and planning but pass multiple hypotheses or occupancy distributions to a planner that scores them against comfort and safety [20,21,57]; others pursue joint training where a differentiable planner shapes the predictor toward plans that are easier to realize [58,59]. Neural state-space models embedded in MPC blend learning's expressiveness with closed-loop guarantees, while planning-as-inference casts decision-making as posterior inference, unifying costs with probabilistic structure [60]. There is also momentum behind world-model approaches that learn simulators for fast rollout and policy improvement.

Two practical gaps remain. First, preserving social multimodality end-to-end: many integrations compress the forecast into a single mean/variance or a soft occupancy map, which washes out intent alternatives crucial for negotiation [20,21]. Second, making the distribution actionable under real-time compute: sample-rich methods expose functional diversity but can be too slow for a tight control loop unless augmented with proposal selection, pruning, and warm-started optimization [22]. These gaps motivate architectures where a learned generator surfaces a compact yet diverse proposal set, and a certified optimizer performs constraint-aware selection and refinement.

3. Method

This paper positions our method squarely at this interface. A Social GAN–style predictor serves as a socially informed proposal generator, yielding a small set of intent-aware trajectories that reflect negotiation outcomes seen in data. Rather than averaging or rasterizing them away, we pass these hypotheses to a constraint-aware MPC that evaluates feasibility, comfort, and risk. Then, we select and refine the best candidate in a closed loop. This preserves mode diversity for high-level intent, yet benefits from MPC's hard guarantees and receding-horizon adaptation. Compared to pipelines that rely on a single forecast, fuse predictions only as soft costs, and adopt heavy sampling at planning time,

our design maintains social consistency while keeping computation predictable via proposal curation and warm-started optimization. The result is a planner robust to intent ambiguity, respectful of safety constraints, and practical under automotive real-time limits. The methods presented in this paper will be detailed in the following subsections: (1) Social GAN for pedestrian trajectory prediction; (2) Model predictive control; and (3) Integration of Social GAN with MPC. The model architecture of this paper is illustrated in Figure 2.

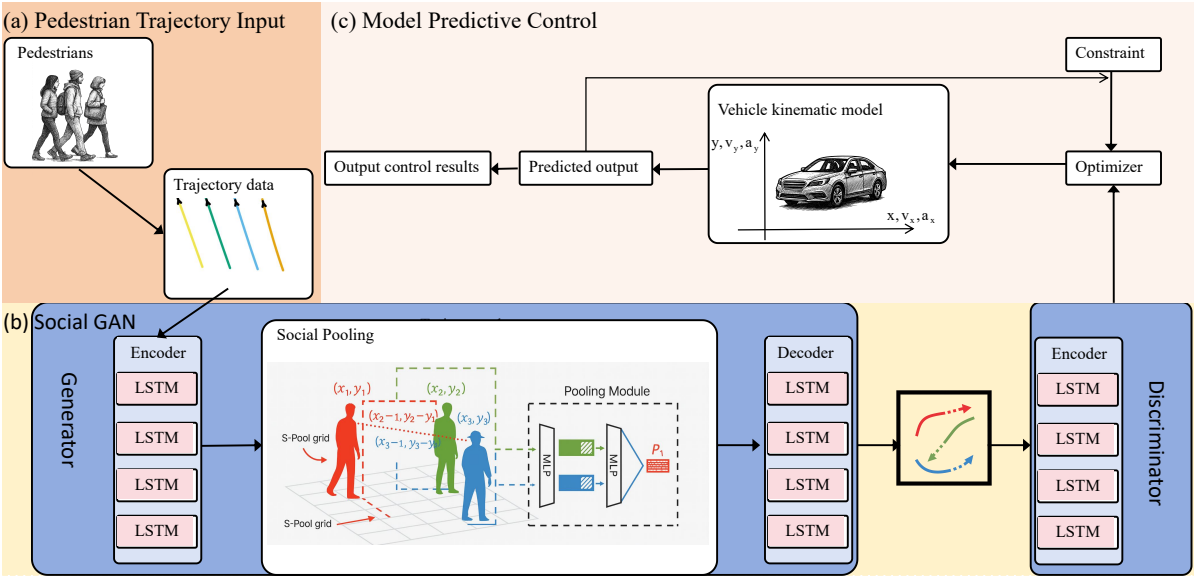

Figure 2. The model architecture of our method.

3.1. Social GAN for pedestrian trajectory prediction

Let $\mathscr{A} = \{1,\ldots,N\}$ be pedestrians visible in a scene. For each agent i, we observe a 2D position sequence over T_{obs} steps, $X_i = \{x_i^t\}_{t=1}^{T_{\text{obs}}}$, and aim to forecast the future T_{pred} steps $Y_i = \{y_i^t\}_{t=T_{\text{obs}}+1}^{T_{\text{obs}}+T_{\text{pred}}}$, where $x_i^t, y_i^t \in \mathbb{R}^2$. We denote $\mathbf{X} = \{X_i\}_{i\in\mathscr{A}}$ and $\mathbf{Y} = \{Y_i\}_{i\in\mathscr{A}}$. Human motion is multi-modal and interaction-dependent; thus the predictor must (i) encode social context and (ii) generate a set of plausible futures.

The generator produces K candidate futures by injecting a latent code $z^{(k)} \sim \mathscr{N}(0,I)$, as shown in Equation (1).

$$h_i^t = \text{LSTM}_{\text{enc}}\left(h_i^{t-1}, \phi\left(x_i^t - x_i^{t-1}\right)\right), \quad t = 2,\ldots,T_{\text{obs}} \tag{1}$$

where $\phi(\cdot)$ is a velocity/relative-displacement embedding. Social interaction at time t is summarized by a permutation-invariant pooling of neighbors' states, as shown in Equation (2).

$$s_i^t = \Psi\left(\left\{g\left(h_j^t - h_i^t, x_j^t - x_i^t\right) \mid j \in \mathscr{N}_i^t\right\}\right) \tag{2}$$

where Ψ an aggregation (e.g., grid-pooling or attention) and g a pairwise feature map. The context vector for agent i is $c_i = \text{Agg}\left(\{[h_i^t, s_i^t]\}_{t=1}^{T_{\text{obs}}}\right)$.

The generator produces K candidate futures by injecting a latent code $z^{(k)} \sim \mathscr{N}(0,I)$, as shown in Equation (3).

$$\hat{Y}_i^{(k)} = \text{G}_\theta\left(c_i, z^{(k)}\right), \quad k = 1,\ldots,K \tag{3}$$

where G_θ is an auto-regressive LSTM decoder, as shown in Equation (4).

$$\hat{y}_i^{t+1,(k)} = \mathrm{Dec}\left(\hat{y}_i^{t,(k)}, c_i, z^{(k)}\right), \quad t = T_{\mathrm{obs}}, \ldots, T_{\mathrm{obs}} + T_{\mathrm{pred}} - 1 \tag{4}$$

where initialized with the last observation $x_i^{T_{\mathrm{obs}}}$. Latent sampling makes $\left\{\hat{Y}_i^{(k)}\right\}$ cover distinct maneuver modes.

A discriminator D_ϕ scores pairs $(\mathbf{X}, \mathbf{Y})$ vs $(\mathbf{X}, \hat{\mathbf{Y}})$. It shares the encoder's social pooling to ensure that social compliance influences real/fake discrimination, as shown in Equation (5).

$$\mathrm{D}_\phi(\mathbf{X}, \mathbf{Y}) \in (0, 1) \tag{5}$$

We combine an adversarial objective with a best-of-K variety loss to encourage both realism and coverage. For generator G and discriminator D, as shown in Equations (6)–(7).

$$\mathscr{L}_{\mathrm{adv}}(G, D) = \mathbb{E}_{(\mathbf{X},\mathbf{Y})}[\log D(\mathbf{X}, \mathbf{Y})] + \mathbb{E}_{\mathbf{X},\mathbf{Z}}[\log(1 - D(\mathbf{X}, \hat{\mathbf{Y}}))] \tag{6}$$

$$\mathscr{L}_{\mathrm{var}}(G) = \sum_{i \in \mathscr{A}} \min_{k \in \{1,\ldots,K\}} \sum_{t=T_{\mathrm{obs}}+1}^{T_{\mathrm{obs}}+T_{\mathrm{pred}}} \left\| y_i^t - \hat{y}_i^{t,(k)} \right\|_1 \tag{7}$$

While the adversarial loss ensures the generated trajectories are socially realistic, it does not inherently encourage diversity among the K samples. To prevent "mode collapse"—where the generator produces multiple similar trajectories that are all plausible but fail to cover the full spectrum of possible futures—we introduce a variety loss Equation (7). Its function is intuitive: for each pedestrian, it looks at all K predicted trajectories and only penalizes the one that is closest to the ground-truth future. This acts as a "best-of-K" training signal, forcing the generator to spread its predictions to cover at least one trajectory close to the actual outcome, thereby explicitly promoting diversity across different latent samples. The overall min–max objective as shown in Equation (8).

$$\min_G \max_D \mathscr{L}_{\mathrm{adv}}(G, D) + \lambda \mathscr{L}_{\mathrm{var}}(G) \tag{8}$$

where $\lambda > 0$ balancing realism and diversity. The variety loss selects, per agent, the closest among K hypotheses, preventing mode collapse while not penalizing diverse but plausible alternatives. The overall schematic diagram of the Social GAN model is shown in Figure 3.

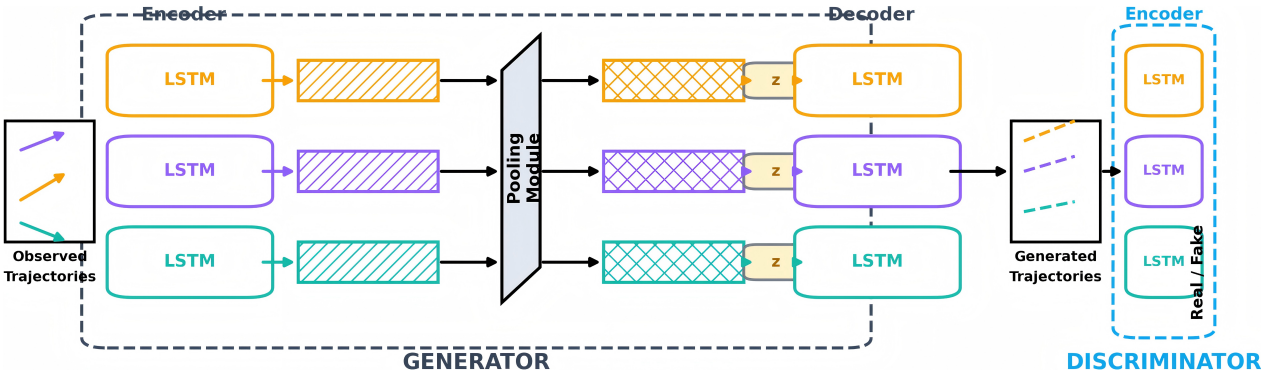

Figure 3. Structure diagram of the Social GAN model.

3.2. Model predictive control

MPC is formulated as a finite-horizon constrained optimization solved at each step; only the first control is applied before the horizon recedes. In our setting, MPC tracks the reference path, respects physical

bounds, and incorporates Social GAN forecasts as dynamic obstacles for safe, smooth planning. The objective function of MPC is typically defined as the weighted sum of the squares of the state error and the control input, as shown in Equation (9).

$$J = \sum_{k=0}^{H} \left\| \mathbf{x}_k - \mathbf{x}_k^{\text{ref}} \right\|_Q^2 + \sum_{k=0}^{H-1} \|\mathbf{u}_k\|_R^2 \tag{9}$$

where $\mathbf{x}_k$ denotes the system state, $\mathbf{u}_k$ denotes the control input, $\mathbf{x}_k^{ref}$ denotes the reference trajectory point, and Q and R are the weighting matrices.

This paper abstracts the vehicle as a point mass model moving in a plane to simplify optimization computation, employing a dual-integral model to describe vehicle dynamics. This model assumes the car can accelerate independently in two orthogonal directions, with motion in each direction governed by two state variables—displacement and velocity—and acceleration control inputs. The system state vector is $\mathbf{x} = [x, ; y, ; v_x, ; v_y]^\top$, where (x, y) represents the vehicle's position in the planar coordinate system and (v_x, v_y) denotes the corresponding velocity components. The control input vector $\mathbf{u} = [a_x, ; a_y]^{\mathrm{T}}$ corresponds to the vehicle's acceleration in the x and y directions. By discretizing the model using a discrete time step Δt, the vehicle state update Equation (10) is obtained.

$$\begin{cases} x_{k+1} = x_k + v_{x,k}\Delta t + \frac{1}{2} a_{x,k} \Delta t^2 \\ y_{k+1} = y_k + v_{y,k}\Delta t + \frac{1}{2} a_{y,k} \Delta t^2 \\ v_{x,k+1} = v_{x,k} + a_{x,k}\Delta t \\ v_{y,k+1} = v_{y,k} + a_{y,k}\Delta t \end{cases} \tag{10}$$

where the subscript k denotes the current discrete time step, and $k+1$ represents the next time step. This paper designs a comprehensive cost function for the vehicle to optimize the MPC path. This cost function comprises a weighted sum of multiple sub-objectives, each corresponding to a specific performance metric. The goal is to quantify deviations in the planned trajectory and control energy consumption, as shown in Equation (11).

$$J_{\text{total}} = \omega_e J_e + \omega_a J_a + \omega_j J_j + \omega_{\text{obs}} J_{\text{obs}} \tag{11}$$

where ω represents the corresponding weighting coefficient, and J represents different cost terms. The path tracking error term is J_e, the acceleration control term is J_a, the acceleration smoothing term is J_j, and the dynamic obstacle avoidance soft constraint term is J_{obs}. The specific cost terms are shown in Equation (12) to Equation (15).

$$J_e = \sum_{k=1}^{H} \left[\left(x_k - x_k^{\text{ref}} \right)^2 + \left(y_k - y_k^{\text{ref}} \right)^2 \right] \tag{12}$$

$$J_a = \sum_{k=0}^{H-1} \left(a_{x,k}^2 + a_{y,k}^2 \right) \tag{13}$$

$$J_j = \sum_{k=0}^{H-2} \left[\left(a_{x,k+1} - a_{x,k} \right)^2 + \left(a_{y,k+1} - a_{y,k} \right)^2 \right] \tag{14}$$

$$J_{\text{obs}} = \sum_{k=1}^{H} \sum_{j=1}^{N_{\text{obs}}} \left\{ \max\left(0, D_{\min} - d_{k,j}\right) \right\}^2 \tag{15}$$

where (x,y) represents the actual position, $\left(x_k^{\text{ref}}, y_k^{\text{ref}}\right)$ represents the desired position, the prediction time window length is H, k denotes the prediction step number, a denotes acceleration, $d_{k,j}$ represents the distance between the obstacle and the vehicle, and $D_{\min}$ denotes the minimum safe distance threshold.

In MPC optimization, in addition to the objective function, a series of constraints must be explicitly imposed to ensure that the resulting control inputs and state sequences are physically feasible and meet safety requirements. Specifically, these include the following types of constraints: dynamic feasibility constraints, physical boundary constraints, and safety constraints. Collectively, these are expressed as Equation (16).

$$\begin{aligned} &\mathbf{X}_{k+1} = A\mathbf{X}_k + B\mathbf{U}_k, \\ &\|\mathbf{v}_k\|_2 \le v_{\max}, \quad \|\mathbf{a}_k\|_2 \le a_{\max} \\ &\left\|\mathbf{p}_k - \mathbf{p}_k^{(i)}\right\|_2 \ge d_{\min}, \quad \forall i,k \end{aligned} \tag{16}$$

where $\mathbf{p}_k^{(i)}$ represents the predicted position of the ith pedestrian at time k, and $v_{\max}$, $a_{\max}$,$d_{\min}$ denote adjustable physical and safety thresholds. In summary, the MPC flowchart is shown in Figure 4.

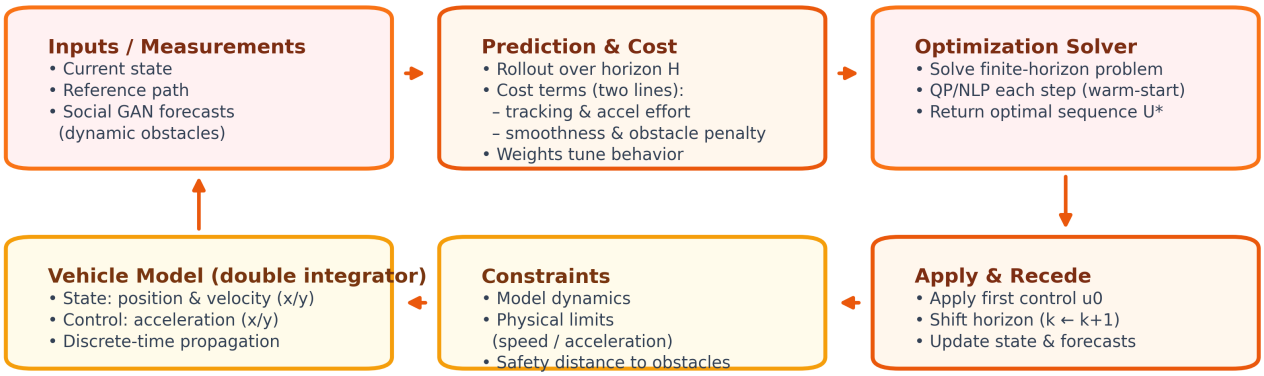

Figure 4. The MPC flowchart.

3.3. Integration of Social GAN with MPC

This paper couples Social GAN's multi-modal forecasts with the MPC in two places already defined in our setup: the objective and the safety constraints. Concretely, we retain the vehicle model, base quadratic tracking/effort terms, and smoothness regularizer in Equations (9)–(14), and we treat Social GAN samples as time-varying dynamic obstacles that enter through the soft avoidance term (Equation (15)) and, when desired, through the distance constraints in Equation (16).

At each control cycle, Social GAN produces K plausible pedestrian futures $\left\{\widehat{Y}^{(k)}\right\}_{k=1}^{K}$ over horizon H. We keep a compact subset for real-time MPC and compute per-sample weights $\{w_k\}$ to emphasize hypotheses most compatible with the recent motion history. This converts the single-mode soft obstacle term in Equation (15) into a prediction-aware composite that preserves social multi-modality without changing the rest of the controller. Let $p_k = [x_k, y_k]^\top$ be the planned ego position. Reusing the hinge-style penalty of Equation (15), we define a weighted sum across the retained hypotheses as Equation (17).

$$J_{\text{obs}}^{\text{SG}} = \sum_k w_k \sum_{k'=1}^{H} \sum_{j=1}^{N_{\text{obs}}} \left[\max\left(0, D_{\min} - \left\|p_{k'} - \hat{y}_{j,k'}^{(k)}\right\|_2\right)\right]^2 \tag{17}$$

where replace J_{obs} in Equation (11) with $J_{\text{obs}}^{\text{SG}}$. For each predicted trajectory hypothesis $\hat{Y}^{(k)}$, its weight w_k is computed as shown in Equation (18).

$$w_k = \frac{\exp\left(-d_{\min}^{(k)}\right)}{\sum_{i=1}^{K} \exp\left(-d_{\min}^{(i)}\right)} \tag{18}$$

where $d_{\min}^{(k)} = \min_{t=1}^{H} \left\| \mathbf{p}_t - \hat{\mathbf{y}}_t^{(k)} \right\|_2$ is the minimum Euclidean distance between predicted trajectory k and the vehicle's planned path over the entire prediction time domain H. No other terms are modified; thus the total cost remains the weighted sum in Equation (11), preserving the design intent and tunability of $\omega_e, \omega_a, \omega_j, \omega_{\text{obs}}$. The whole integrated framework is detailed in Algorithm 1.

Algorithm 1 Social GAN–Informed MPC with Multi-Modal Dynamic-Obstacle Aggregation

Inputs: horizon H, step Δt, weights $(\omega_e, \omega_a, \omega_j, \omega_{\text{obs}})$, bounds $v_{\max}, a_{\max}, D_{\min}$; reference path $\{(x_k^{\text{ref}}, y_k^{\text{ref}})\}_{k=1}^{H}$; number of predictions K; retain budget $M \le K$.

State: ego state $\mathbf{x}$; warm-start sequence $\mathbf{U}^{\text{ws}}$ (optional).

while task not finished **do**
  Observe pedestrian histories $\mathbf{X} = \{X_i\}_{i \in \mathscr{A}}$ over T_{obs}.
  Encode social context to obtain $\{c_i\}$ using (1)–(2).
  Sample $\{z^{(k)}\}_{k=1}^{K} \sim \mathscr{N}(0, I)$ and generate K futures $\widehat{\mathbf{Y}}^{(k)} = \{\hat{Y}_i^{(k)}\}_{i \in \mathscr{A}}$ via (3)–(4).
  Compute compatibility weights $\{w_k\}_{k=1}^{K}$ from recent motion; normalize so $\sum_{k=1}^{K} w_k = 1$.
  Select top-M hypotheses $\mathscr{K} \subseteq \{1, \ldots, K\}$ by w_k (keeps runtime predictable).
  Let $p_k = [x_k, y_k]^\top$ denote the planned ego position along the horizon.
  Build $J_{\text{obs}}^{\text{SG}}$ as in (17) using $\{w_k\}_{k \in \mathscr{K}}$ and $\{\widehat{\mathbf{Y}}^{(k)}\}_{k \in \mathscr{K}}$.
  Formulate MPC: minimize $J_{\text{total}} = \omega_e J_e + \omega_a J_a + \omega_j J_j + \omega_{\text{obs}} J_{\text{obs}}^{\text{SG}}$ with J_e, J_a, J_j from (12)–(14), subject to dynamics (10) (or the state-space form in (16)), speed/accel bounds $\|\mathbf{v}_k\|_2 \le v_{\max}$, $\|\mathbf{a}_k\|_2 \le a_{\max}$, and (optional) hard distances $\|\mathbf{p}_k - \mathbf{p}_k^{(i)}\|_2 \ge d_{\min}$.
  Warm-start with $\mathbf{U}^{\text{ws}}$ if available.
  Solve for $\mathbf{U}^{\star} = \{\mathbf{u}_0^{\star}, \ldots, \mathbf{u}_{H-1}^{\star}\}$ and $\{\mathbf{x}_k^{\star}\}_{k=1}^{H}$.
  Apply $\mathbf{u}_0^{\star}$; set $\mathbf{U}^{\text{ws}} \leftarrow [\mathbf{u}_1^{\star}, \ldots, \mathbf{u}_{H-1}^{\star}, \mathbf{0}]$.
  Advance time, update $\mathbf{x}$, refresh observations/predictions.
end while

4. Experimental setup and results

4.1. Dataset, experimental setup, and evaluation metrics

This paper selects ETH [61]/UCY [62] as the dataset, comprising the ETH BIWI pedestrian dataset and the UCY exemplar crowd dataset. The two datasets encompass five widely used scenarios: ETH and HOTEL (derived from ETH), and UNIV, ZARA1, and ZARA2 (derived from UCY). ETH was recorded in Zurich from a bird's-eye view over building entrances and sidewalks, capturing numerous real pedestrian tracks and avoidance behaviors; UCY covers open-space campus and street areas with group behaviors such as joining, merging, and yielding. Trajectories are typically derived from 25-fps videos. Owing to

rich social interactions and occlusions, ETH/UCY has become a canonical resource for evaluating social consistency and cross-scene generalization.

The experimental environment for this paper includes hardware and hyperparameter settings, as detailed in Table 1.

Table 1. Hyperparameter settings.

Hyperparameter	Value
Learning rate	0.0015
Epochs	300
Batch size	4
Observation time	3.2 s (8 frames)
Prediction time	4.8 s (12 frames)
Sequence Length	8 s (20 frames)
Sampling number	20
Optimizer	Adam
Sampling period	0.4 s
Prediction Step Size	8
Control horizon	8
Maximum Number of Iterations for Optimal Response	5
Control Convergence Criteria	0.001

The evaluation metrics in this paper assess performance from two aspects: trajectory prediction and control performance. The evaluation metrics for trajectory prediction are ADE and FDE, as shown in Equations (19)–(20).

$$\mathrm{ADE} = \frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\hat{y}_i^t - y_i^t\right\|_2 \tag{19}$$

$$\mathrm{FDE} = \frac{1}{N}\sum_{i=1}^{N}\left\|\hat{y}_i^T - y_i^T\right\|_2 \tag{20}$$

where T denotes the predicted time step, N denotes the total number of samples, and $\hat{y}_i^t$ and y_i^t represent the predicted position and actual position of the ith sample at frame t, respectively. The performance evaluation metrics for control are Time to Goal, Path Efficiency, Safety Violations, Average Acceleration, and Average Jerk, as shown in Equations (21)–(25).

$$T_{\text{goal}} := \inf\left\{t \geq 0 \mid \left\|\mathbf{s}(t) - \mathbf{s}_{\text{goal}}\right\|_2 < \varepsilon, \|\mathbf{v}(t)\|_2 < \delta\right\} \tag{21}$$

where $\mathbf{s}_k/\mathbf{s}(t)$ is the position vector, $\mathbf{v}_k/\mathbf{v}(t)$ is the velocity vector, $\varepsilon, \delta > 0$ is the distance and velocity threshold, and T_{goal} is the time at which the target is first approached while simultaneously meeting the velocity threshold, as shown in Equation (22).

$$\eta = \frac{L_{\text{optimal}}}{L_{\text{actual}}} \tag{22}$$

where $L_{\text{optimal}} = \left\|\mathbf{s}_0 - \mathbf{s}_{\text{goal}}\right\|$ represents the shortest path and $L_{\text{actual}} = \sum_{i=1}^{n}\|\mathbf{s}_i - \mathbf{s}_{i-1}\|$ represents the cumulative actual driving distance, as shown in Equation (23).

$$N_{\text{violation}} = \sum_{k=1}^{K} \mathbb{I}\left(\min_{j} \left\| \mathbf{s}_k - \mathbf{p}_k^{(j)} \right\| < d_{\min}\right) \tag{23}$$

where $\mathbf{S}_k$ denotes the robot's position at time k, $\mathbf{p}_k^{(j)}$ represents the position of the jth pedestrian at that time, and $\mathbb{I}(\cdot)$ is an indicator function, as shown in Equation (24).

$$\bar{a} = \frac{1}{K} \sum_{k=1}^{K} \|\mathbf{a}_k\|_2 \tag{24}$$

where $\mathbf{a}_k$ is the acceleration vector for step k, and K is the total number of control steps, as shown in Equation (25).

$$\bar{j} = \frac{1}{K-1} \sum_{k=2}^{K} \frac{\|\mathbf{a}_k - \mathbf{a}_{k-1}\|_2}{\tau} \tag{25}$$

where τ is the sampling period.

4.2. Experimental results and analysis

To further demonstrate the superiority of the proposed method, experiments were conducted in this section, and the relevant experimental results were analyzed. The comparative results of the trajectory prediction models are shown in Table 2.

Table 2. Comparative experimental results of trajectory prediction models (prediction performance).

Method (ADE/FDE)	ETH	HOTEL	UNIV	ZARA01	ZARA02	Average
LSTM	0.70/1.45	0.55/1.77	0.36/0.77	0.25/0.53	0.31/0.65	0.43/1.03
Social LSTM [63]	0.73/1.48	0.49/1.01	0.41/0.84	0.27/0.56	0.33/0.70	0.45/0.92
SGAN [23]	0.60/1.19	0.52/1.02	0.44/0.84	0.22/0.43	0.29/0.58	0.41/0.81

Comparison of ADE/FDE Performance Between Social GAN and other trajectory prediction methods on the ETH/UCY Dataset (Lower values indicate fewer errors). This paper presents ADE and FDE results for Social GAN and Social-LSTM across five sub-scenarios: ETH, HOTEL, UNIV, ZARA1, and ZARA2. As shown in Table 2, Social GAN performs better than other methods, delivering more diverse predicted trajectories.

To evaluate the effectiveness of the proposed trajectory-prediction and control strategy, we conduct two experiments in a dynamic crowd environment, comparing: (i) a simple reactive obstacle-avoidance baseline (emergency-stop policy) and (ii) the proposed Social GAN-assisted MPC controller. Both experiments share identical start/goal configurations and crowd dynamics; Social GAN predicts pedestrian trajectories, while motion decisions are executed by either the emergency-stop mechanism or the MPC controller. We report performance using six quantitative metrics: arrival time, final path efficiency, average minimum distance to pedestrians, number of safety violations, mean acceleration, and jerk. The metric definitions are summarized in Table 3, and the comparative results are presented in Figures 5 and 6.

In Figures 5 and 6, the red dot trajectory represents the robot, moving from the lower-left starting point toward the upper-right. The green pentagram marks the robot's target point. Trajectories of other colors indicate the movement paths of several pedestrians in the scene, showing multiple pedestrians

crossing near the robot's route. The robot must maintain a safe distance from pedestrians without halting forward motion. Both control strategies leverage pedestrian trajectory predictions provided by Social GAN to anticipate and avoid collisions.

Table 3. Comparative experimental results of our method (control performance).

Performance Metrics	**Simple reactive obstacle-avoidance**	**Social GAN + MPC**
Arrival time	13.6 s	14.4 s
Final path efficiency	99.91%	96.31%
Average minimum distance to pedestrians	0.93 m	0.94 m
Number of safety violations (< 0.1 m)	1	0
Mean acceleration (m/s^2)	0.04	0.57
Jerk (m/s^3)	0.11	1.34

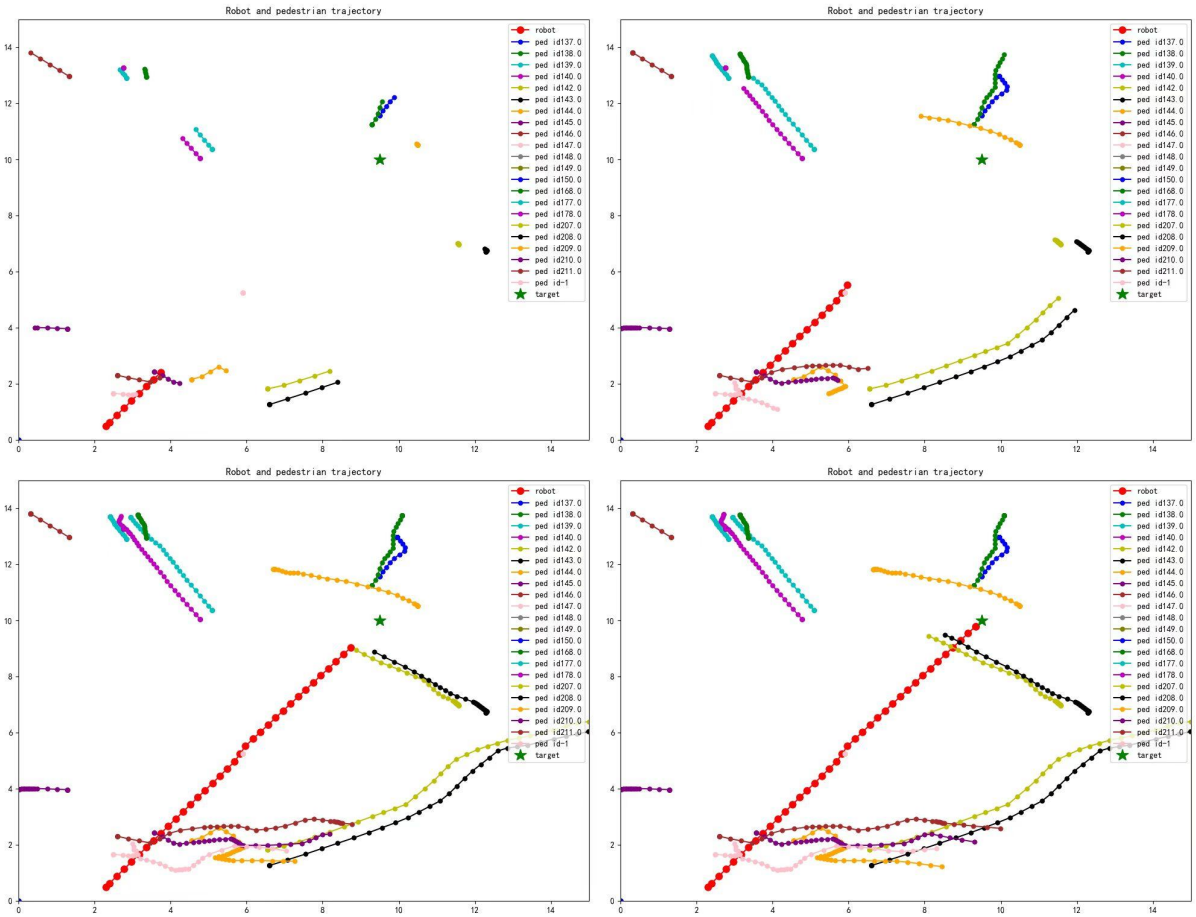

Figure 5. Visualization of simple reactive obstacle avoidance.

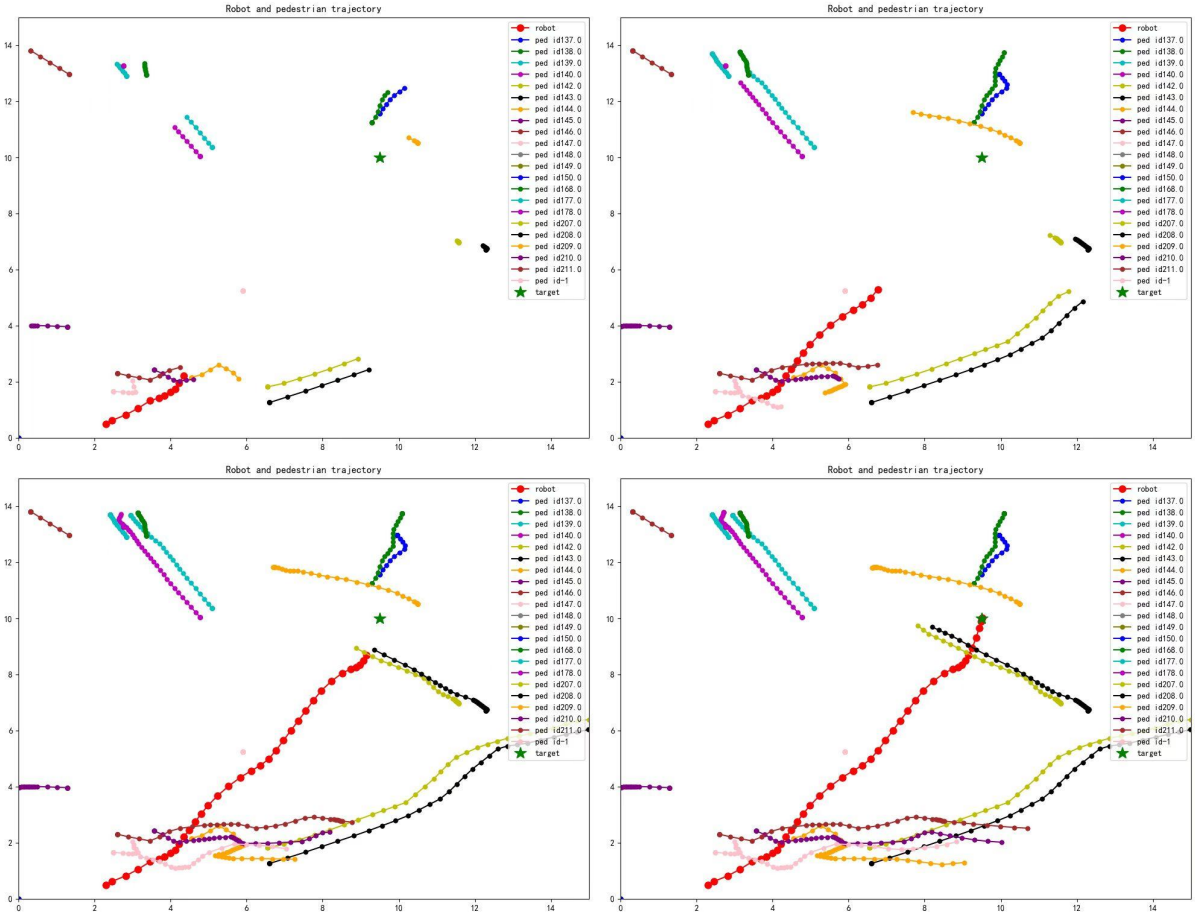

Figure 6. Visualization of Social GAN + MPC.

In dynamic-crowd simulations, the Social GAN-assisted MPC achieves zero safety violations and a slightly larger average minimum pedestrian clearance (0.94 m *vs.* 0.93 m) than a simple emergency-stop baseline. This enhanced safety is achieved at the cost of marginally longer travel time (14.4 s *vs.* 13.6 s) and reduced path efficiency (96.31% *vs.* 99.91%) due to risk-aware detours. Crucially, the controller's proactive nature is reflected in the elevated mean acceleration and jerk (0.57 m/s^2 and 1.34 m/s^3 *vs.* 0.04 m/s^2 and 0.11 m/s^3). The significant increase in jerk, which quantifies the rate of change of acceleration, is a direct result of the MPC continuously refining the vehicle's trajectory in response to multiple, evolving pedestrian predictions. Unlike the static emergency-stop policy, our method makes smoother, earlier adjustments to anticipated threats, which nevertheless involve more frequent changes in acceleration than simply maintaining a constant velocity or executing a single, hard brake.

Overall, the Social GAN+MPC controller trades negligible delay and modest path elongation for materially improved safety and feasibility, aligning with zero-collision engineering requirements; further weight tuning can reduce jerk without compromising avoidance performance.

4.3. Real-time verification results and analysis

To verify whether the proposed framework satisfies the predictable computation characteristics required for real-time control, we conducted a detailed analysis of the latency of the system's key components. The results in Figures 7–9 clearly confirm that the system can meet the real-time control demands of

autonomous driving while maintaining computational predictability.

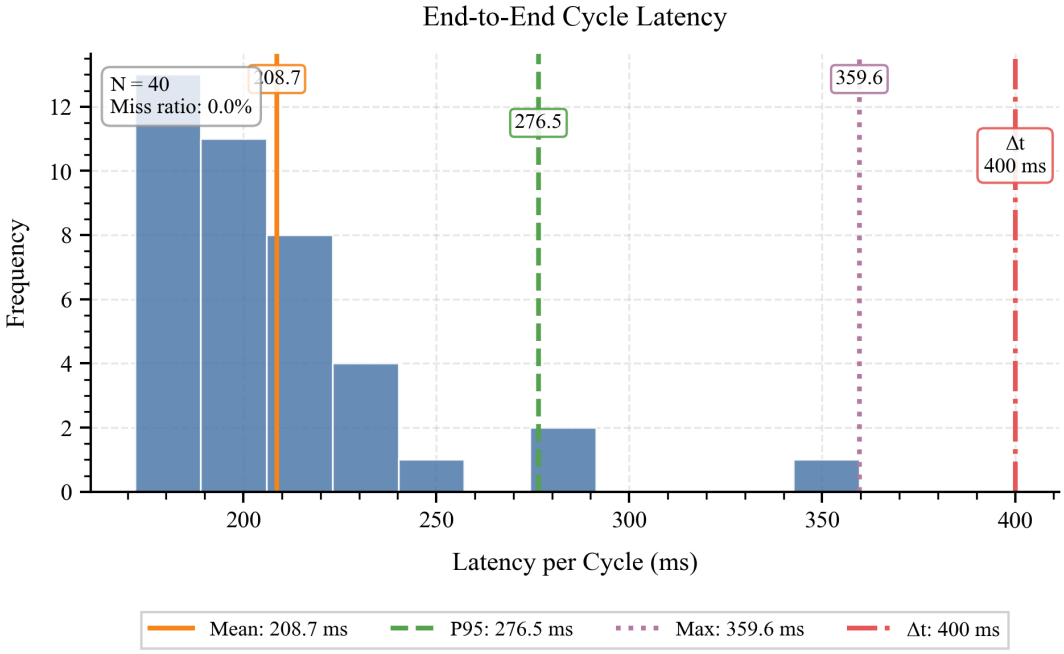

Figure 7. End-to-end prediction-to-control pipeline latency distribution.

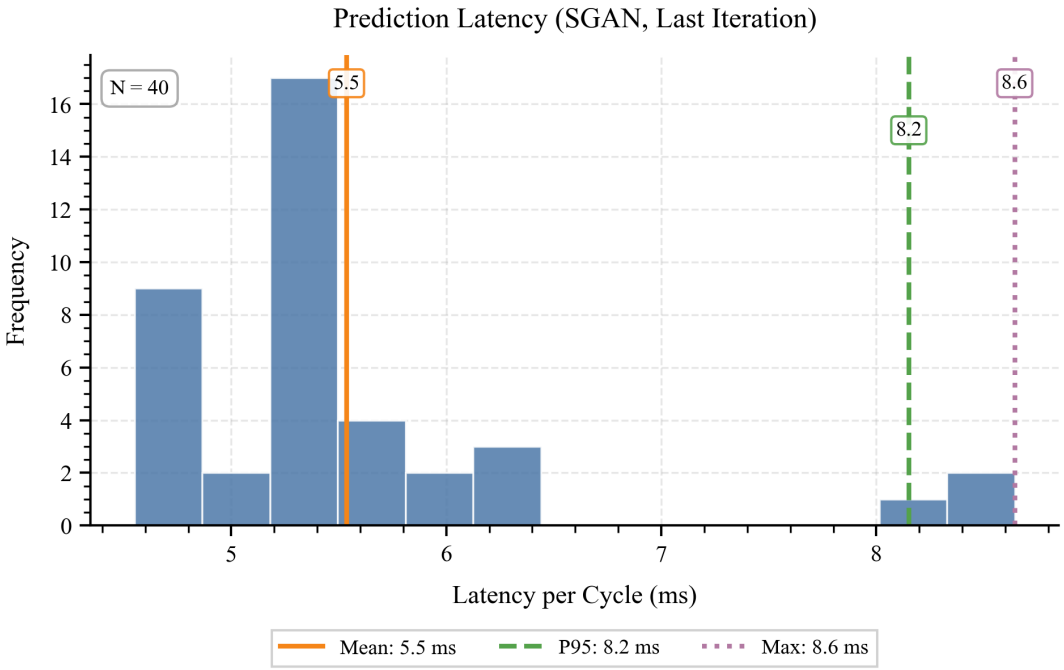

Figure 8. Social GAN multimodal trajectory prediction module latency distribution.

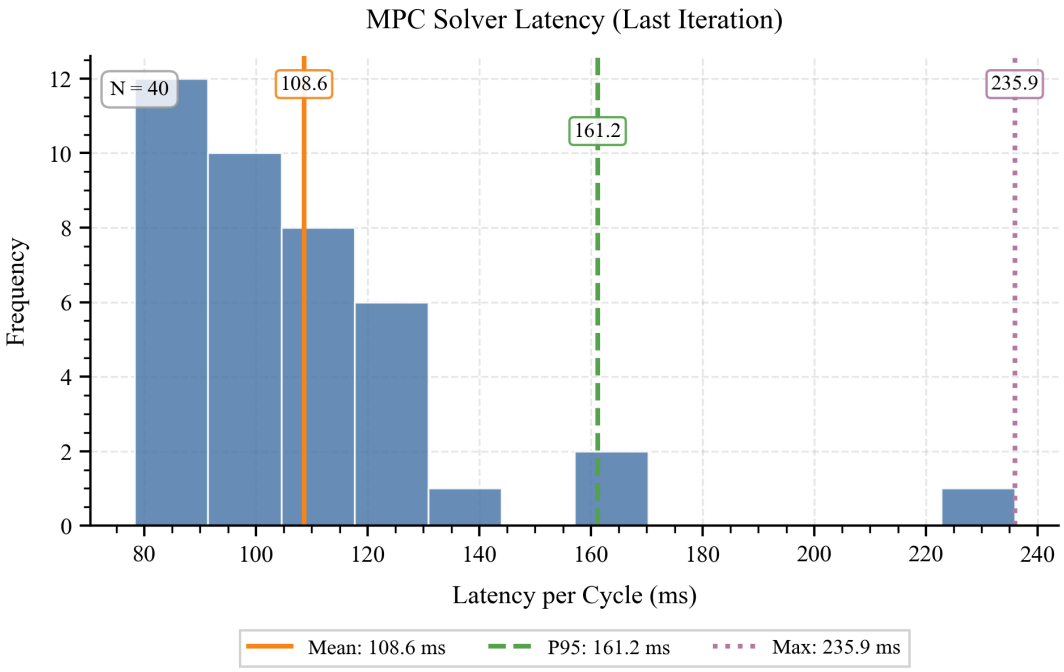

Figure 9. MPC solver single-iteration optimization latency distribution.

The end-to-end latency analysis (Figure 7) reveals an average latency of 208.7 milliseconds across 40 complete control cycles, with a 95th percentile (P95) latency of 276.5 milliseconds and a maximum latency of 359.6 milliseconds. Crucially, all cycles completed within the 400-millisecond deadline with a 0% failure rate. This outcome demonstrates that the entire prediction-control pipeline maintains high reliability under stringent temporal constraints, providing a robust foundation for real-time deployment.

Further decomposition of latency sources reveals that the Social GAN prediction module (Figure 8) exhibits extremely low computational overhead, with an average latency of only 5.5 milliseconds and a P95 latency of 8.2 milliseconds. This minimal and stable latency proves that generating multimodal trajectory predictions does not become a real-time control bottleneck, enabling the integration of complex deep learning predictors within the control loop.

As the primary computational component, the MPC solver (Figure 9) exhibits an average latency of 108.6 milliseconds and a P95 latency of 161.2 milliseconds. This performance indicates that the optimization solver can complete within reasonable control cycles even when considering complex scenarios with multiple pedestrian trajectory hypotheses. Combined with the prediction module's latency, the total latency of the core algorithm (prediction + planning) averages 114.1 milliseconds with a P95 of 169.4 milliseconds, entirely suitable for typical autonomous driving control cycles (typically 100–200 milliseconds).

The compactness of these latency distributions—where all components' P95 and maximum values fall within reasonable ranges—further validates the system's computational predictability. This predictability is critical for safety-critical autonomous driving systems, as it ensures controllers respond promptly across diverse scenarios without incurring unacceptable computational delays.

5. Conclusion

This paper has introduced a unified prediction-to-control pipeline that effectively bridges a critical gap in autonomous driving: integrating socially aware, multimodal trajectory prediction with real-time, safety-certifiable control. Our key contribution lies in demonstrating that preserving intent diversity from a predictor like Social GAN and directly embedding it into an MPC's optimization process is feasible and essential for achieving robust zero-collision performance in dynamic environments.

The experimental findings validate the core premise of our approach. The achievement of zero safety violations in dense simulations underscores that reasoning over multiple plausible futures is a decisive advantage over single-mode or reactive strategies. The associated marginal increases in travel time and control effort are not merely trade-offs, but evidence of a proactive and strategically cautious planning philosophy. The fact that these metrics remained within acceptable comfort bounds confirms that our method delivers substantial safety improvements without compromising passenger comfort or real-time feasibility.

The implications of this work extend beyond the specific Social GAN and MPC models. It provides a generalizable framework for fusing learning-based multimodal prediction with optimization-based control, highlighting the importance of computationally efficient proposal curation to make such integration practical. This moves the field away from brittle, single-future planning and towards a more robust paradigm where autonomous vehicles can confidently navigate the inherent uncertainty of human behavior.

Looking ahead, this framework opens several promising directions [64,65]. We will refine the controller's comfort awareness through jerk-aware regularization and more sophisticated cost functions. Furthermore, tightening the coupling between prediction and planning—through uncertainty-aware weighting and even joint training—promises to yield more coherent and efficient behaviors. Finally, incorporating formal risk-sensitive constraints (e.g., chance constraints, Control Barrier Functions) and validating the system in hardware-in-the-loop and real-world experiments are critical next steps toward deploying this technology in safety-critical autonomous systems. While the proposed framework demonstrates robust performance in simulation, we note the theoretical challenge of ensuring recursive feasibility under multimodal predictions. Future work will focus on integrating stochastic MPC with chance constraints or control barrier functions to provide rigorous safety certificates under prediction uncertainty.

Data availability statement

All data used in this paper are from open-source datasets. You can access it from this link: https://github.com/cwang-nus/ETH-UCY-datasets?tab=readme-ov-file.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China under Grants No. 62573292, 62206199, 62476192 and U2141234; the National Science and Technology Major Project under Grant No. 2022ZD0119900; the National Key Laboratory of Marine Engine Science and

Technology under Grant No. LAB-2024-04-WD; the Young Elite Scientist Sponsorship Program under Grant No. YESS20220409; the Natural Science Foundation of Tianjin under Grant No. 23JCQNJC02010; and the Hainan Province Science and Technology Special Fund under Grant No. ZDYF2024GXJS003. The authors also would like to thank the editors and anonymous reviewers for their time and effort in evaluating this article and for the constructive comments for the improvement of its presentation and quality.

Authors' contribution

Conceptualization, Z.J. and H.Z.; methodology, H.Z.; software, Z.J.; validation, H.Z., Z.J. and X.H.; formal analysis, R.L.; investigation, Z.J.; resources, W.Z.; data curation, Z.J.; writing—original draft preparation, Z.J., H.Z. and X.H.; writing—review and editing, R.L. and W.Z.; visualization, Z.J. and X.H.; supervision, R.L.; project administration, W.Z.; funding acquisition, R.L. All authors have read and agreed to the published version of the manuscript.

Conflicts of interests

The authors declare no conflict of interest.

References

[1] Jiang Z, Guo Y, Jiang K, Hu M, Zhu Z. Optimization of intelligent plant cultivation robot system in object detection. *IEEE Sens. J.* 2021, 21(17):19279–19288.

[2] Jiang Z, Zhang R, Guo Y, Hu M, He L, *et al*. Noise interference reduction in vision module of intelligent plant cultivation robot using better cycle GAN. *IEEE Sens. J.* 2022, 22(11):11045–11055.

[3] Liu R, Hu X, Jiang Z, Wang J, Zhang W. Survey on heterogeneous aquatic robot systems: communication, perception, navigation, control, decision-making and energy management. *Robot Learn.* 2025, 2(1):1–61.

[4] Korta P, Kurramsetty VK, Iyer LV, Kar NC. Discontinuous control strategy for loss reduction in wound-field synchronous machine based electric vehicle powertrain system. *IEEE J. Emerging Sel. Top. Ind. Electron.* 2025, 6(3):900–909.

[5] Li D, Zhang J, Liu G. Autonomous driving decision algorithm for complex multi-vehicle interactions: an efficient approach based on global sorting and local gaming. *IEEE Trans. Intell. Transp. Syst.* 2024, 25(7):6927–6937.

[6] Wang X, Li K, Chehri A. Multi-sensor fusion technology for 3D object detection in autonomous driving: a review. *IEEE Trans. Intell. Transp. Syst.* 2024, 25(2):1148–1165.

[7] Lee D, Seong H, Kang G, Han S, Shim DH, *et al*. Design, field evaluation, and traffic analysis of a competitive autonomous driving model in a congested environment. *IEEE Trans. Intell. Transp. Syst.* 2024, 25(8):9482–9497.

[8] Fu Z, Jiang K, Xie C, Xu Y, Huang J, *et al*. Summary and reflections on pedestrian trajectory prediction in the field of autonomous driving. *IEEE Trans. Intell. Veh.* 2024.

[9] Golchoubian M, Ghafurian M, Dautenhahn K, Azad NL. Pedestrian trajectory prediction in pedestrian-vehicle mixed environments: a systematic review. *IEEE Trans. Intell. Transp. Syst.* 2023, 24(11):11544–11567.

[10] Yang B, Yan S, Wang Z, Nakano K. Prediction based trajectory planning for safe interactions between autonomous vehicles and moving pedestrians in shared spaces. *IEEE Trans. Intell. Transp. Syst.* 2023, 24(10):10513–10524.

[11] Jiang Z, Qin C, Yang R, Shi B, Alsaadi FE, *et al* . Social entropy informer: a multi-scale model-data dual-driven approach for pedestrian trajectory prediction. *IEEE Trans. Intell. Transp. Syst.* 2025, 26(10):16438–16453.

[12] Jiang Z, Ma Y, Shi B, Lu X, Xing J, *et al*. Social NSTransformers: low-quality pedestrian trajectory prediction. *IEEE Transactions on Artificial Intelligence* 2024, 5(11):5575–5588.

[13] Jiang Z, Yang R, Ma Y, Qin C, Chen X, *et al*. Social informer: pedestrian trajectory prediction by informer with adaptive trajectory probability region optimization. *IEEE Trans. Cybern.* 2025.

[14] Zhang D, Liang J, Lu S, Guo K, Wang Q, *et al*. PEP: policy-embedded trajectory planning for autonomous driving. *IEEE Rob. Autom. Lett.* 2024, 9(12):11361–11368.

[15] Zhou X, Zhao W, Wang A, Wang C, Zheng S. Spatiotemporal attention-based pedestrian trajectory prediction considering traffic-actor interaction. *IEEE Trans. Veh. Technol.* 2023, 72(1):297–311.

[16] Bhaskara R, Viswanath H, Bera A. Trajectory prediction for robot navigation using flow-guided Markov neural operator. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, May 13–17, 2024, pp. 15209–15216.

[17] Liang R, Li Y, Zhou J, Li X. STGlow: a flow-based generative framework with dual-Graphormer for pedestrian trajectory prediction. *IEEE Trans. Neural Networks Learn. Syst.* 2024, 35(11):16504–16517.

[18] Liu Y, Li B, Wang X, Sammut C, Yao L. Attention-aware social graph transformer networks for stochastic trajectory prediction. *IEEE Trans. Knowl. Data Eng.* 2024, 36(11):5633–5646.

[19] Meng Q, Guo H, Li J, Liu J, Liu Z, *et al*. Map search-based vehicle trajectory prediction conditions for lane lines with heterogeneous interaction in complex urban traffic. *IEEE Trans. Ind. Inf.* 2025, 21(9):6958–6967.

[20] Phan-Minh T, Grigore EC, Boulton FA, Beijbom O, Wolff EM. CoverNet: multimodal behavior prediction using trajectory sets. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, June 14–19, 2020, pp. 14062–14071.

[21] Mahjourian R, Kim J, Chai Y, Tan M, Sapp B, *et al*. Occupancy flow fields for motion forecasting in autonomous driving. *IEEE Rob. Autom. Lett.* 2022, 7(2):5639–5646.

[22] Liao B, Chen S, Yin H, Jiang B, Wang C, *et al*. DiffusionDrive: truncated diffusion model for end-to-end autonomous driving. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, USA, June 11–15, 2025, pp. 12037–12047.

[23] Gupta A, Johnson J, Fei-Fei L, Savarese S, Alahi A. Social GAN: socially acceptable trajectories with generative adversarial networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, USA, June 18–22, 2018, pp. 2255–2264.

[24] Thompson M, Dallas J, Goh JYM, Balachandran A. Adaptive nonlinear model predictive control: maximizing tire force and obstacle avoidance in autonomous vehicles. *IEEE Trans. Field Rob.* 2024,

1:318–331.

[25] Jia Z, Zhang K, Shi Y, Zhang W. Safety-preserving Lyapunov-based model predictive rendezvous control for heterogeneous marine vehicles subject to external disturbances. *IEEE Trans. Cybern.* 2024, 54(9):5244–5256.

[26] Shi L, Wang L, Long C, Zhou S, Tang W, *et al*. Representing multimodal behaviors with mean location for pedestrian trajectory prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* 2023, 45(9):11184–11202.

[27] Li W, Zhang Y, Li L, Lv Y, Wang M. A pedestrian trajectory prediction model for right-turn unsignalized intersections based on game theory. *IEEE Trans. Intell. Transp. Syst.* 2024, 25(8):9643–9658.

[28] Lv K, Yuan L, Ni X. Learning autoencoder diffusion models of pedestrian group relationships for multimodal trajectory prediction. *IEEE Trans. Instrum. Meas.* 2024, 73:1–12.

[29] Zhu P, Zhao S, Deng H, Han F. Attentive radiate graph for pedestrian trajectory prediction in disconnected manifolds. *IEEE Trans. Intell. Transp. Syst.* 2025, 26(6):7755–7769.

[30] Hu C, Niu R, Lin Y, Yang B, Chen H, *et al*. Probabilistic trajectory prediction of vulnerable road user using multimodal inputs. *IEEE Trans. Intell. Transp. Syst.* 2025, 26(2):2679–2689.

[31] Li S, Zou W, Gao J, Yin Y, Kim D, *et al*. Fast online computation of MPC-based integrated decision control for autonomous vehicles. *IEEE Trans. Intell. Veh.* 2024, 9(11):7059–7069.

[32] Wei H, Shi Y. Mpc-based motion planning and control enables smarter and safer autonomous marine vehicles: perspectives and a tutorial survey. *IEEE/CAA J. Autom. Sin.* 2023, 10(1):8–24.

[33] Zhang H, Liu Y, Zhao W, Hu C, Zhao J. Human–machine shared control for steer-by-wire vehicles using improved reinforcement learning-based MPC. *IEEE Trans. Intell. Transp. Syst.* 2025, 26(8):12688–12700.

[34] Askari I, Vaziri A, Tu X, Zeng S, Fang H. Model predictive inferential control of neural state-space models for autonomous vehicle motion planning. *IEEE Trans. Rob.* 2025, 41:3202–3222.

[35] Adelirad M, Afzalian AA. A computationally efficient trajectory prediction in MPC for piecewise affine systems. *IEEE Trans. Autom. Control* 2024, 69(9):6215–6221.

[36] Pang S, Cao J, Jian M, Lai J, Yan Z. BR-GAN: a pedestrian trajectory prediction model combined with behavior recognition. *IEEE Trans. Intell. Transp. Syst.* 2022, 23(12):24609–24620.

[37] Dendorfer P, Elflein S, Leal-Taixé L. MG-GAN: a multi-generator model preventing out-of-distribution samples in pedestrian trajectory prediction. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, Canada, October 11–17, 2021, pp. 13138–13147.

[38] Sharma N, Dhiman C, Indu S. Cross-modal pedestrian behavior prediction: a dual-task approach With progressive denoising attention and CVAE. *IEEE Trans. Intell. Transp. Syst.* 2025, 26(10):17110–17120.

[39] Bhattacharyya A, Reino DO, Fritz M, Schiele B. Euro-PVI: pedestrian vehicle interactions in dense urban centers. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, USA, June 19–25, 2021, pp. 6408–6417.

[40] Liu Y, Zhang Y, Li K, Qiao Y, Worrall S, *et al*. Knowledge-aware graph transformer for pedestrian trajectory prediction. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, Bilbao, Spain, September 24–28, 2023, pp. 4360–4366.

[41] Kim S, Chi H, Lim H, Ramani K, Kim J, *et al*. Higher-order relational reasoning for pedestrian

trajectory prediction. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, June 17–21, 2024, pp. 15251–15260.

[42] Feng A, Qiu R, Wang J, Gong J, Yi Y, *et al*. Multimodal forward generation transformer network for inconspicuous pedestrian trajectory prediction. *IEEE Rob. Autom. Lett.* 2024, 9(3):2224–2231.

[43] Zhang Z, Zhou J, Liu S, Xiao B. Completed interaction networks for pedestrian trajectory prediction. *IEEE Trans. Multimedia* 2025, 27:5119–5129.

[44] Ivanovic B, Pavone M. Rethinking trajectory forecasting evaluation. *arXiv* 2021, arXiv:2107.10297.

[45] Saadatnejad S, Bahari M, Khorsandi P, Saneian M, Moosavi-Dezfooli S, *et al*. Are socially-aware trajectory prediction models really socially-aware? *Transp. Res. Part C Emerg. Technol.* 2022, 141:103705.

[46] Cao C, Chen X, Wang J, Song Q, Tan R, *et al*. CCTR: calibrating trajectory prediction for uncertainty-aware motion planning in autonomous driving. In *2024 Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, Vancouver, Canada, February 20–27, 2024, pp. 20949–20957.

[47] Bouzidi M, Schlauch C, Scheuerer N, Yao Y, Klein N, *et al*. Closing the loop: motion prediction models beyond open-loop benchmarks. *arXiv* 2025, arXiv:2505.05638.

[48] Rhinehart N, Mcallister R, Kitani K, Levine S. PRECOG: PREdiction conditioned on goals in visual multi-agent settings. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Republic of Korea, October 27–November 2, 2019, pp. 2821–2830.

[49] Mayne DQ. Model predictive control: recent developments and future promise. *Automatica* 2014, 50(12):2967–2986.

[50] Paden B, Čáp M, Yong SZ, Yershov D, Frazzoli E. A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Trans. Intell. Veh.* 2016, 1(1):33–55.

[51] Pinneri C, Sawant S, Blaes S, Achterhold J, Stueckler J, *et al*. Sample-efficient cross-entropy method for real-time planning. In *Proceedings of the 2020 Conference on Robot Learning, PMLR*, London, UK, November 8–11, 2021, pp. 1049–1065,

[52] Ames AD, Coogan S, Egerstedt M, Notomista G, Sreenath K, *et al*. Control barrier functions: theory and applications. In *2019 18th European Control Conference (ECC)*, Naples, Italy, June 25–28, 2019, pp. 3420–3431.

[53] Wenzel M, Din ED, Zimmer M, Benigni A. Gaussian process supported stochastic MPC for distribution grids. *IEEE Open J. Control Syst.* 2025, 4:332–348.

[54] Xu M, Chen Y, Hu M, Yang Y. Risk-aware informative path planning for information gthering of a 3D surface. *IEEE Trans. Autom. Sci. Eng.* 2025, 22:20584–20595.

[55] Jiao Y, Yang K, Song D. Federated distributionally robust optimization with non-convex objectives: algorithm and analysis. *IEEE Trans. Mob. Comput.* 2025.

[56] Zhang X, Zeinali S, Schildbach G. Interaction-aware traffic prediction and scenario-based model predictive control for autonomous vehicles on highways. *IEEE Trans. Control Syst. Technol.* 2025, 33(4):1235–1245.

[57] McAllister R, Wulfe B, Mercat J, Ellis L, Levine S, *et al*. Control-aware prediction objectives for autonomous driving. In *2022 International Conference on Robotics and Automation (ICRA)*, Philadelphia, USA, May 23–27, 2022, pp. 1–8.

[58] Huang Z, Liu H, Wu J, Lv C. Differentiable integrated motion prediction and planning with

learnable cost function for autonomous driving. *IEEE Trans. Neural Networks Learn. Syst.* 2024, 35(11):15222–15236.

[59] Huang Z, Karkus P, Ivanovic B, Chen Y, Pavone M, *et al*. DTPP: differentiable joint conditional prediction and cost evaluation for tree policy planning in autonomous driving. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, May 13–17, 2024, pp. 6806–6812.

[60] Askari I, Vaziri A, Tu X, Zeng S, Fang H. Model predictive inferential control of neural state-space models for autonomous vehicle motion planning. *IEEE Trans. Rob.* 2025, 41:3202–3222.

[61] Pellegrini S, Ess A, Van Gool L. Improving data association by joint modeling of pedestrian trajectories and groupings. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision (ECCV)*, Heraklion, Greece, September 5–11, 2010, pp. 452–465.

[62] Lerner A, Chrysanthou Y, Lischinski D. Crowds by example. *Comput. Graphics Forum* 2007, 26(3):655–664.

[63] Alahi A, Goel K, Ramanathan V, Robicquet A, Fei-Fei L, *et al*. Social LSTM: human trajectory prediction in crowded spaces. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, June 27–July 1, 2016, pp. 961–971.

[64] Jin B, Cruz L, Gonçalves N. Pseudo RGB-D face recognition. *IEEE Sens. J.* 2022, 22(22):21780–21794.

[65] Jin B, Gonçalves N, Cruz L, Medvedev I, Yu Y, *et al*. Simulated multimodal deep facial diagnosis. *Expert Syst. Appl.* 2024, 252:123881.